Journal of Clinical Gerontology & Geriatrics 2 (2011) 83–87

Contents lists available at ScienceDirect

Journal of Clinical Gerontology & Geriatrics

journal homepage: www.e-jcgg.com

Original article

A brief questionnaire is able to measure population physical activity levels accurately: A comparative validation study

Stephanie Webster, B Phty (Hons), Asad Khan, PhD, Jennifer Catherine Nitz, PhD *

Division of Physiotherapy, School of Health & Rehabilitation Sciences, The University of Queensland, St Lucia, Brisbane, Queensland 4072, Australia

ARTICLE INFO

Article history:
Received 16 March 2011
Received in revised form
4 May 2011
Accepted 14 June 2011

Keywords:
Physical activity level
Questionnaires
Validation

ABSTRACT

Background: Categorizing individuals on the basis of physical activity (PA) level has both epidemiological utility and public health implications. Questionnaires provide a combination of accuracy and practicality when assessing population PA, but their characteristics, including summarization of raw data and scaling methods, can affect utility in large-scale studies. The aim of this study was to ascertain the concurrent validity of a brief six-point scale for the categorical scaling of PA levels. The effect of participant characteristics on the scales was also assessed.

Methods: A comparative observational study was used. One hundred participants aged 18–80 years completed the valid and reliable International Physical Activity Questionnaire (IPAQ) and the Human Activity Profile (HAP) and a six-point scale. Quadratic weighted kappa, Spearman's rho, and Fisher's exact tests were applied to compare the questionnaire results for concurrent validity. The effect of participant characteristics on PA level was analyzed using Fisher's exact test or analysis of variance.

Results: Kappa values comparing the six-point scale with the IPAQ ($\kappa = 0.46$) and HAP ($\kappa = 0.57$) showed fair to good agreement. Analysis using Fisher's exact test illustrated a significant association between the six-point scale and both the HAP ($p < 0.000$) and IPAQ ($p < 0.000$). Spearman's rho correlation values between the six-point scale and the IPAQ and the HAP were 0.49 and 0.69, respectively. Scores on all questionnaires were significantly related to age ($p < 0.05$). Marital status and number of comorbidities were related to HAP and six-point scale scores ($p < 0.05$). Occupation was significantly related to HAP but not the other two scales ($p < 0.000$).

Conclusion: The six-point scale provides a quick measurement of usual amount and intensity of PA that a person performs. This validation study found significant associations and a moderate strength of agreement when compared with the IPAQ and HAP. The six-point scale has the potential for utility in large-scale epidemiological studies as a quick method for scaling individuals by PA level.

1. Introduction

The importance of physical activity in maintaining health throughout life is well recognized with a worldwide consensus that an increase in physical activity (PA) level is needed for people of all ages.[1,2] With greater emphasis on the relationship of PA with health, there is an emerging need for accurate, valid, and practical methods of assessing PA. One of the benefits of classifying individuals on the basis of PA level is that categorical distinctions of low, moderate, and high have both epidemiological utility and public health implications in terms of promotion and intervention strategies.[3] However, accurate assessment of PA levels in the general population is difficult and can be time consuming for both participants and researchers. When considering the levels of PA in older people, the burden of assessment needs to be minimal. A number of methods for measuring PA exist, from precise laboratory techniques to single-item questionnaires. Direct measurement of PA by accelerometer[4] and energy expenditure by doubly labeled water technique[5] are currently the objective measures of choice. Doubly labeled water technique is not practical for large-scale studies because of high cost, participant burden, and the intrusiveness of the technique.[5] Similarly, accelerometers have seen limited use because of high cost, participant burden, and staff demands. Furthermore, they tend to overpredict the cost of walking and are unable to detect arm movements, external work performed in pushing or lifting, stair climbing, or carrying one's body weight uphill.[6] Most frequently, PA has been assessed using a variety of

* Corresponding author. Division of Physiotherapy, School of Health & Rehabilitation Sciences, The University of Queensland, St Lucia, Brisbane, Queensland 4072, Australia.
E-mail address: j.nitz@uq.edu.au (J. C. Nitz).

2210-8335
doi:10.1016/j.jcgg.2011.06.003

questionnaires. Questionnaires provide a combination of accuracy and practical application when assessing population activity levels.[7] However, questionnaire characteristics, including summarization of raw data and scaling methods, differ and can affect feasibility in large-scale studies where PA may not be the main exposure variable.

Two commonly used questionnaires are the International Physical Activity Questionnaire (IPAQ) and the Human Activity Profile (HAP). The IPAQ enables collection of detailed information within domains of household and yard work activities, occupational activity, transportation, and leisure-time PA. It has been shown to be a valid and reliable method for monitoring population levels of PA in diverse settings.[8] The HAP measures activity levels based on the calculation of an adjusted activity score (AAS) representing an estimate of an individual's average energy expenditure.[9] Collection of data from the IPAQ and the HAP places considerable burden on individuals and staff. They are both lengthy questionnaires to complete and require additional time to review and code, increasing the cost of studies. When applied to older people, there is often the need for assistance in completing the questionnaires. When this time or funding is not available, a shorter questionnaire to categorize PA is necessary.

Several studies have recently reported the development of brief PA assessment tools. Wendel-Vos et al[10] reported the reproducibility and validity of the Short Questionnaire to Assess Health-enhancing Physical Activity (SQUASH). This tool is able to classify individuals into three different intensity categories—light, moderate, or vigorous. Completing the SQUASH takes about 3–5 minutes; however, scoring is complex involving calculations for each separate question using intensity scores based on metabolic equivalent (MET) values. A MET is the ratio of the rate of energy expended during an activity to the rate of energy expended at rest. The quality of the SQUASH validity study was limited by the small sample size of 50, resulting in an inability to generalize findings and less-precise estimates of validity. Moreover, the weighted kappa value assessing the overall relative validity of the SQUASH was low ($\kappa = 0.30$).

Matthews et al[11] developed a short PA recall questionnaire to determine if a person was meeting current PA recommendations. The Short Telephone-administered Activity Recall questionnaire[11] used information gained from respondents' participation in moderate- and vigorous-intensity activity to classify the participants as physically inactive, insufficiently active, or meet current recommendations. The validity coefficients ($r = 0.10–0.40$) and kappa values ($\kappa = 0.14–0.43$) for the Short Telephone-administered Activity Recall questionnaire were low, and the limited participant age range (20–48 years) excludes the use of this tool with an older population.

The Stanford Brief Activity Survey is a two-item tool developed to classify individuals into five PA categories and to detect if the individual is meeting the recommended PA guidelines.[12] The use of a select age range (60–69 years) limits the ability to generalize the applicability of the Stanford Brief Activity Survey until further validation in other age groups is performed.

A brief six-point scale was devised by Hirvensalo et al[13] measuring the level of PA and its intensity. In this scale, activity is scored from 1 to 6 with several descriptors referring to perspiring and depth of breathing. Significant positive associations between the number of sweat episodes per week and total weekly leisure-time energy expenditure, as assessed by the Harvard Alumni Activity Survey,[14] have been shown. This six-point scale[13] appears to be the simplest method of categorizing PA levels that takes these indicators of PA into account. However, to our knowledge, there have been no direct comparative studies analyzing the level of agreement between this scale and other commonly used measures.

The research questions this study aimed to answer are as follows: Is the scaling method used in the six-point scale[13] comparable to MET-min/wk categories of the IPAQ and the categorical scale of the HAP for a general population?; Does the six-point scale[13] have a validity similar to those of more complex scales?; and do the results suggest, for studies gathering multiple data, that the six-point scale might be seen as a choice option?

2. Methods

2.1. Participants

A volunteer sample was recruited from the community through verbal invitation to participate at the university, local hospital, and at local businesses. A minimum population size of 100 has been recommended for a validation study of PA questionnaires.[15] To enable the results of this work to be used across an aging population, the age range for inclusion was 18–80 years. The aim was to recruit equal numbers of men and women. Participants were selected independent of race or comorbidities. Exclusion criteria were men or women not living independently in the community or unable to give informed consent.

All participants were given an information sheet, and the study aims and procedures were also explained verbally. All participants gave written informed consent for participation. The study protocol was approved by the University Ethical Review Committee and undertaken in accordance with the Declaration of Helsinki.

2.2. Protocol

Participants were asked to fill out three questionnaires: the IPAQ, HAP, and the six-point scale.[13] The HAP collected demographic information, including age, gender, marital status, comorbidities, occupation, current smoking status, and highest school grade completed. Questionnaires were filled out independently by the participants who had access to the research assistant for clarification if required.

2.3. Measurements

Marital status was categorized as single, married, or divorced/separated. Those participants who were widowed were classified as divorced/separated. Comorbidities were dichotomized as either yes, where one or more were specified, or no. Participants were asked to specify their occupation. Based on occupational job titles, participants were categorized using the Physical Demands Strength Rating.[16] Participant's occupation was classified as sedentary, light, moderate, heavy, or very heavy, reflecting the overall strength requirement of the job.

The strength rating reflects the workers' involvement in standing, walking, and sitting, as well as the intensity and duration of lifting, carrying, pushing, and pulling. Retirees and housewives were not recognized as having occupations by the HAP or the Physical Demands Strength Rating reference. It has been suggested by Cabrera de León et al[17] that people undertaking less than 25–30 minutes of PA daily or less than 10% of leisure-time energy expenditure of four or less MET[18] were sedentary. We considered this definition would fit retirees and housewives in our study; hence, they were classified as sedentary.

2.3.1. International Physical Activity Questionnaire

The IPAQ uses MET energy expenditure estimates from the compendium of PAs to code PA by intensity.[19] Participants completed the IPAQ long version, which assessed duration, frequency, and intensity of physical activities in the past week. The

total scores for all walking and moderate and vigorous PAs were computed using MET values and formulae from the guidelines for data processing and analysis of the IPAQ.[20] These values were then summed to form a continuous measure of total PA MET-min/wk score. Using this score and other specified criteria (number of days of moderate and/or vigorous and/or walking, and number of minutes per day), participants were classified categorically as "low," "moderate," or "high." The questionnaire includes an additional question that measures time spent sitting as an indicator of sedentary activity; however, this is not included as part of the summary PA score and was, therefore, not used. Test-retest reliability data for the long IPAQ questionnaire has been shown to be very good (pooled $r = 0.81$ across data from 12 countries), and criterion validity against accelerometers has shown a fair to moderate agreement (pooled $r = 0.33$).[8]

2.3.2. HAP questionnaire

The HAP consists of a list of 94 activities ranked in ascending order of level of energy required to perform each activity. A wide variety of activities are represented, including self-care tasks, personal/household work, entertainment/social activities, and independent exercise pursuits. Participants completed the HAP assigning each activity to one of the three categorical options: 1 = still doing this activity, 2 = have stopped doing this activity, or 3 = never did this activity. The HAP is calculated by subtracting the number of less-demanding activities a participant has stopped performing from a calculated maximal activity score. HAP activity scores calculated include the maximal activity score (MAS) and the AAS. The MAS is the number identifying the activity with the highest oxygen consumption requirement that the participant still performs. The AAS is the difference between the MAS and the number of less-demanding activities the participant has stopped performing. The developers of the HAP considered the AAS to be a more stable estimate of the individual's daily activities and the best estimate of the responder's average energy expenditure.[9] Bennell et al[21] found that, compared with the MAS, the AAS appeared to be the most useful parameter for detecting change and, perhaps, the most clinically relevant. Each parameter was scored from 0 to 94, with higher scores representing greater PA. PA was classified based on the AAS as 1 = impaired (<53); 2 = moderately active (53–74); or 3 = active (>74).

2.3.3. The six-point scale[13]

Participants completed the six-point scale.[13] In this scale, activity is scored as 1 = moving only for necessary chores; 2 = walking or other outdoor activities one or two times per week; 3 = walking or other outdoor activities several times per week; 4 = exercising one or two times per week to the point of perspiring and heavy breathing; 5 = exercising several times per week to the point of perspiring and heavy breathing; or 6 = keep-fit heavy exercise or competitive sport several times per week. Hirvensalo et al[13]qualified the scores greater than 3 by having participants identify the PA or sport they were undertaking and frequency of participation. Hirvensalo et al[13] categorized the types of sport or activity into brisk, which included skiing or swimming, and moderate, which included walking for fitness or calisthenics. When examples of sport or exercise were inappropriate to our participant cohort because of climate, perspiring and panting were considered appropriate to be used as a guide to describe the level of exertion when applied to the sport or activity identified by the participant. This method of determining exertion has been used previously[22] when indoor gymnasium-based activities, such as aerobic, yoga, and Pilates classes were included owing to the popularity of these activities locally. To enable comparison between the three scales, we collapsed the six-point scale into three categorical levels: low, moderate, and high. We proposed low (scores 1–2, which is consistent with the category of sedentary used by the developers of the six-point scale[13]); moderate (scores 3–4); and high (scores 5–6) categories. This classification best approximated the three-level categorical scales used in both the HAP and IPAQ and, hence, enabled more direct comparisons.

2.4. Data analysis

SPSS for Windows, version 17 (SPSS Inc., Chicago, IL, USA), was used for most of the analytical procedures. Cohen's kappa is commonly used to determine the percent agreement beyond that expected by chance; however, it distinguishes only between agreement and disagreement.[23] To allow for differential weightings of disagreement, quadratic weighted kappa was calculated. As per a previous study,[11] guidelines of Landis and Koch[24] were followed for describing the strength of agreement. Kappa values of 0.39 or less indicate "poor" agreement, values between 0.40 and 0.75 indicate "fair to good," and values greater than 0.75 indicate "excellent" agreement. Spearman's correlation coefficients were calculated among the categorical scores on the six-point scale,[13] the classifications based on AAS in the HAP, and categorical classifications based on MET min/wk in the IPAQ. These were calculated so that comparisons could be undertaken between our results and those of previous studies. Overall relative validity was investigated by using Fisher's exact test.

Analysis of variance was performed to determine if there were age differences between activity-level categories in each of the three scales. Fisher's exact test was used to analyze the relationship of the remaining demographic characteristics with the PA-level categories of the three scales. Statistical significance was defined as $p < 0.05$ for all statistical tests.

3. Results

A total of 100 community-dwelling adults participated in the study. Collectively, the participants represented a wide range of ages, levels of education, occupations, and activity levels. The mean age of the participants was 49.3 years and ranged from 18 years to 80 years. Seventy-eight percent of the participants were women. Most of the participants were married (61%), and 72% had completed high school. A total of 10% were smokers, and 25% of the study population had one or more comorbidities. Population comorbidities included arthritis, chronic lower back pain, pulmonary fibrosis, hypertension, cardiac disease, diabetes, hearing impairment, depression, asthma, Parkinson's disease, bronchiectasis, high cholesterol, and multiple sclerosis. No participants in this study had an occupation classified as heavy or very heavy. Approximately half (49%) of the participants had an occupation categorized as sedentary.

Age was found to be significantly related to the PA level in each of the three reported scales. The relationship between the remaining demographic characteristics and activity levels of the three scales are presented in Table 1 along with participant representation in each PA category. Gender and smoking status had a nonsignificant effect on categorical activity levels in the three scales. Marital status was found to have a significant effect on the activity-level scaling in the HAP. The presence of one or more comorbidities was found to have a significant effect on the activity-level scaling in the HAP and the six-point scale.[13] A significant relationship was found between occupational physical demands and PA level in the HAP. Although participants with greater occupational physical demands were found more likely to be classified as physically active, those with sedentary occupations were not found more likely to be physically inactive. Only 5% of the

Table 1
Effect of participant characteristics on frequency (%) of category representation by physical activity measure for 100 participants (78% female)

Characteristics	IPAQ			HAP (AAS)			6-Point scale		
	Low	Mod	High	Impaired	Mod	High	Low	Mod	High
Gender									
Male	4.5	45.5	50	0	45.5	54.5	27.25	27.25	45.5
Female	11.5	33.3	55.2	6.4	46.2	47.4	29.4	32.1	38.5
Marital status									
Single	8	23	30	3	32	26	18	22	21
Married	2	8	18	1	6	21	5	6	17
Divorced/separated	0	5	6	1	8	2*	6	3	2**
Comorbidities									
0	5	25	45	1	28	46	16	22	37
≥1	5	11	9	4	18	3***	13	9	3***
Smoking status									
Nonsmoker	10	34	46	5	43	42	26	28	36
Smoker	0	2	8	0	3	7	3	3	4
Occupation, activity level									
Sedentary	6	25	18	5	29	15	16	19	14
Light	4	8	18	0	14	16	11	6	13
Moderate	0	3	18	0	3	18***	2	6	13

* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.000$.
AAS = adjusted activity score; HAP = Human Activity Profile; IPAQ = International Physical Activity Questionnaire; Mod = moderate.

participants with sedentary occupations were classified as impaired (HAP < 53) with the HAP and 10% with the IPAQ. The six-point scale[13] found that nearly one-third (29%) of the participants with sedentary occupations were classified as having a low level of PA.

Relative validity comparisons between the categorized scale[13] and the HAP and IPAQ are presented in Table 2. Weighted kappa analysis showed a "fair to good" agreement with both the IPAQ and HAP. A stronger agreement is noted between the six-point scale[13] and the HAP. When the weighted Kappa analysis was applied between the IPAQ and HAP, agreement was at 0.38, signifying a poor one. Analysis using Fisher's exact test illustrated a significant association between the six-point scale[13] and both the HAP and IPAQ. Modest but significant correlations between the six-point scale[13] and the reference questionnaires were found using Spearman's rho correlations.

4. Discussion

Our aim was to determine the concurrent validity of the six-point scale[13] for categorically assigning individuals based on the general levels of PA. Results from this study showed significant associations and a moderate strength of agreement between the six-point scale[13] and the two reference questionnaires.

For our participant cohort, we found a poor agreement between the IPAQ and HAP scales. This finding would suggest that each of these scales was measuring different aspects of PA. Because our findings showed a fair to good agreement between the six-point scale[13] (categorized to three levels) and the IPAQ and HAP scales, respectively, it suggests that the six-point scale[13] more broadly described the PA level of our participants. This initial evaluation suggests that, for epidemiological studies gathering multiple data and from participant groups comprising older people, the six-point scale[13] might be a valid choice for categorizing participants by PA level.

Schmidt and Steindorf[25] highlighted the difficulty of defining acceptable agreement between two PA questionnaires. Criteria regarding what constitutes good agreement do not exist. Validation studies of most questionnaires primarily base their conclusions on Pearson's or Spearman's correlation coefficients. They are measures of association but not of agreement, and caution should be taken when concluding acceptable validity even if there is a high correlation.[25] Spearman's rho correlations found in the present study lie within the range of correlation coefficients found in relation to other direct and indirect measures of PA.[7,10] The strength of our study is that we further calculated quadratic weighted kappa values as a measure of agreement. Also a large age range was included in our study, validating its utility for a large percentage of the adult population.

Most population-based studies have focused on leisure-time PA.[26] Excluding measures of occupational PA may result in significant underestimations of PA in those employed in physically demanding occupations.[27] Individuals with sedentary jobs likely perform the bulk of their PA during leisure time, whereas those with physically demanding occupations may be more active during work.[28] Our results support these conclusions showing that those participants with more physically demanding occupations were more likely to be classified as physically active. Those with sedentary occupations were infrequently classified as inactive and, hence, were likely to perform PA within other domains. Although the six-point scale[13] was not as sensitive as the reference questionnaires with these associations, categorical levels from the scale were found to have a significant association with occupational physical demands. The estimate of PA obtained using the six-point scale[13]

Table 2
Quadratic weighted kappa, Fisher's exact test, and Spearman's rho correlation analyses between the six-point scale[13] and the reference questionnaires

Measures	Kappa		Fisher's exact test (p)	Spearman's rho (r)
	Observed kappa (SE)	95% CI (lower, upper)		
HAP	0.57 (0.08)	0.41–0.73	0.000	0.69
IPAQ	0.46 (0.10)	0.26–0.66	0.000	0.49

CI = confidence interval; HAP = Human Activity Profile; IPAQ = International Physical Activity Questionnaire; SE = standard error.

might reflect a more global PA estimate by recognizing occupational activity level in addition to leisure-time activity when choosing PA category. Further research is needed to assess the inclusion of occupation-related PA in the six-point scale[13] if this deviation from protocol was adopted.

Classifications of occupational physical demands in our study were estimates based on occupational titles with no individual measures made. This could increase the risk of misclassification because different persons with the same job title may have a large variability in the PA performed.[28] A more accurate classification has been proposed by using a combination of self-report and objective measurements.[29]

Misclassification might have occurred when retirees and housewives in our study were classified as sedentary. Our finding that the sedentary group was infrequently classified as participating in low levels of activity suggests that our assumption for classifying retirees and housewives was inaccurate, and these individuals likely had diverse levels of PA participation.

Self-administered questionnaires as criterion measures have been previously used to validate other questionnaires.[12] Both reference questionnaires in this study have previously been validated using objective measures that directly measured movement or PA-related energy expenditure. The authors cannot completely exclude the possibility that measurement error between the assessments exists and, therefore, potential overestimation of agreement.

Some limitations should be considered when evaluating the validity of the six-point scale.[13] Originally, the scale was used with older people. For this study, a broad adult age range participated with no adverse effect, though this aspect was not addressed. Also, the sample size was relatively small. Small samples can result in limited ability to generalize results to other populations and can lead to less-precise estimates of validity.[30] Correlation coefficients become less precise when based on smaller samples because of an increased standard error.[31] However, validity in our study was not judged solely on the basis of correlations but on the methods more sensitive to the level of individual errors in reporting. Future validation should use both self-report and objective measures of activity in parallel analyses. This would aid clarification of average energy expenditure equivalent to each activity category, facilitating both interpretation and comparisons among studies. Additional research to extend this initial study is needed to determine the sensitivity of the six-point scale[13] to detect meaningful changes in habitual PA.

In conclusion, in this initial validation study, we found significant, though moderate, strength of agreement between the six-point scale[13] and the IPAQ and HAP. The six-point scale[13] appears to more broadly describe the PA level of our participants than either the IPAQ or the HAP and provides a quick measurement of usual amount and intensity of PA that a person performs. The six-point scale[13] has the potential for utility in large-scale epidemiological studies as a quick method for scaling individuals by PA level.

References

1. Littman A, Kristal A, White E. Effects of physical activity intensity, frequency, and activity type on 10-y weight change in middle-aged men and women. *Int J Obes* 2005;**29**:524–33.
2. Vuillemin A, Boini S, Bertrais S, Tessier S, Oppert J-M, Hercberg S, et al. Leisure time physical activity and health-related quality of life. *Prev Med* 2005;**41**: 562–9.
3. Caspersen C, Powell K, Christenson G. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Rep* 1985;**100**:126–31.
4. Troiano R, Berriggan D, Dodd K, Masse L, Tilert T, McDowell M. Physical activity in the United States by accelerometer. *Med Sci Sports Exerc* 2008;**40**:181–8.
5. Irwin M, Ainsworth B, Conway J. Estimation of energy expenditure from physical activity measures: determinants of accuracy. *Obes Res* 2001;**9**:517–25.
6. Bassett D, Ainsworth B, Swartz A, Strath S, O'Brein W, King G. Validity of four motion sensors in measuring moderate-intensity physical activity. *Med Sci Sports Exerc* 2000;**32**(Suppl 9):s471–80.
7. Miller D, Freedson P, Kline G. Comparison of activity levels using the Caltrac accelerometer and five questionnaires. *Med Sci Sports Exerc* 1994;**26**:376–82.
8. Craig C, Marshall A, Sjostrom M, Bauman A, Booth M, Ainsworth B, et al. International Physical Activity Questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc* 2003;**35**:1381–95.
9. Daughton D, Fix A, Kass I, Bell C, Patil K. Maximum oxygen consumption and the ADAPT Quality-of-Life Scale. *Arch Phys Med Rehabil* 1982;**63**:620–2.
10. Wendel-Vos G, Schuit A, Saris W, Kromhout D. Reproducibility and relative validity of the Short Questionnaire to Assess Health-enhancing Physical Activity. *J Clin Epidemiol* 2003;**56**:1163–9.
11. Matthews C, Ainsworth B, Hanby C, Pate R, Addy C, Freedson P. Development and testing of a Short Physical Activity Recall Questionnaire. *Med Sci Sports Exerc* 2005;**37**:986–94.
12. Taylor-Piliae R, Norton R, Haskell W, Mahbouda M, Fair J, Iribarren C. Validation of a new brief physical activity survey among men and women aged 60-69 years. *Am J Epidemiol* 2006;**164**:598–606.
13. Hirvensalo M, Rantanen T, Heikkinen E. Mobility difficulties and physical activity as predictors of mortality and loss of independence in the community-living older population. *J Am Geriatr Soc* 2000;**48**:493–8.
14. Washburn R, Goldfield S, Smith K, McKinlay J. The validity of self-reported exercise-induced sweating as a measure of physical activity. *Am J Epidemiol* 1990;**132**:107–13.
15. Pols M, Peeters P, Kemper H, Grobbee D. Methodological aspects of physical activity assessment in epidemiological studies. *Eur J Epidemiol* 1998;**4**: 63–70.
16. United States Department of Labor Employment and Training Administration. *Dictionary of occupational titles*. Washington DC: United States Department of Labor Employment and Training Administration; 1991.
17. Cabrera de León A, de la Rodríguez-Pérez M, Rodríguez-Benjumeda L, Anía-Lafuente B, Brito-Díaz B, de Fuentes M, et al. Sedentary lifestyle: physical activity duration versus percentage of energy expenditure. *Rev Esp Cardiol* 2007;**60**:244–50.
18. Varo JJ, Martinez-Gonzalez MA, de Irala-Estevez J, Kearney J, Gibney M, Martinez JA. Distribution and determinants of sedentary lifestyles in the European Union. *Int J Epidemiol* 2003;**32**:138–46.
19. Ainsworth B, Haskell M, Whitt M, Irwin M, Swartz A, Strath S, et al. Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc* 2000;**32**:S498–516.
20. Guidelines for data processing and analysis of the international physical activity questionnaire (IPAQ)-short and long forms, 2005, http://www.ipaq.ki.se/scoring.htm [accessed 12.09.2009].
21. Bennell K, Hinman R, Crossley K, Metcalf B, Buchbinder R, Green S. Is the Human Activity Profile a useful measure in people with knee osteoarthritis? *J Rehabil Res Dev* 2004;**41**:621–30.
22. Nitz J, Low Choy NL. Changes in activity level in women aged 40-80 years. *Climacteric* 2007;**10**:408–15.
23. Shuster C. A note on the interpretation of weighted kappa and its relations to other rater agreement statistics for metric scales. *Ed Psych Meas* 2004;**64**: 243–53.
24. Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–74.
25. Schmidt M, Steindorf K. Statistical methods for the validation of questionnaires: discrepancy between theory and practice. *Methods Inf Med* 2006;**45**: 409–13.
26. Sallis J, Saelens B. Assessment of physical activity by self-report: status, limitations, and future directions. *Res Q Exerc Sport* 2000;**71**:1–14.
27. Steele R, Mummery K. Occupational physical activity across occupational categories. *J Sci Med Sport* 2003;**6**:398–407.
28. Dorn J, Cerny F, Epstein L, Naughton J, Vena J, Winkelstein W, et al. Work and leisure time physical activity and mortality in men and women from a general population sample. *Ann Epidemiol* 1999;**9**:366–73.
29. Bassett D. Validity and reliability issues in objective monitoring of physical activity. *Res Quart Exer Sports* 2007;**1**(2 Suppl):S30–6.
30. Neilson H, Robson P, Friedenreich C, Csizmadi I. Estimating activity energy expenditure: how valid are physical activity questionnaires? *Am J Clin Nutr* 2008;**87**:279–91.
31. Herbert J, Miller D. The inappropriateness of conventional use of the correlation coefficient in assessing validity and reliability of dietary assessment methods. *Eur J Epidemiol* 1991;**17**:339–43.